Comparison of additive image fusion vs. feature-level image fusion techniques for enhanced night driving

Edward J. Bender [a], Colin E. Reese [a], Gooitzen S. van der Wal [b]

[a] US Army CERDEC Night Vision & Electronic Sensors Directorate, Fort Belvoir, VA 22060
[b] Sarnoff Corporation, 201 Washington Road, Princeton, NJ 08540

ABSTRACT

The Night Vision & Electronic Sensors Directorate (NVESD) has conducted a series of image fusion evaluations under the Head-Tracked Vision System (HTVS) program. The HTVS is a driving system for both wheeled and tracked military vehicles, wherein dual-waveband sensors are directed in a more natural head-slewed imaging mode. The HTVS consists of thermal and image-intensified TV sensors, a high-speed gimbal, a head-mounted display, and a head tracker. A series of NVESD field tests over the past two years has investigated the degree to which additive (A+B) image fusion of these sensors enhances overall driving performance. Additive fusion employs a single (but user adjustable) fractional weighting for all the features of each sensor's image. More recently, NVESD and Sarnoff Corporation have begun a cooperative effort to evaluate and refine Sarnoff's "feature-level" multi-resolution (pyramid) algorithms for image fusion. This approach employs digital processing techniques to select at each image point only the sensor with the strongest features, and to utilize only those features to reconstruct the fused video image. This selection process is performed simultaneously at multiple scales of the image, which are combined to form the reconstructed fused image. All image fusion techniques attempt to combine the "best of both sensors" in a single image. Typically, thermal sensors are better for detecting military threats and targets, while image-intensified sensors provide more natural scene cues and detect cultural lighting. This investigation will address the differences between additive fusion and feature-level image fusion techniques for enhancing the driver's overall situational awareness.

Keywords: Image Fusion, Image-Intensified TV, Thermal Sensor, Helmet-Mounted Display (HMD), Head-Tracking, Head-Tracked Vision System (HTVS), and nighttime driving.

1. INTRODUCTION

For years, US military vehicle drivers have relied upon image intensifiers such as the AN/PVS-7 Night Vision Goggle (wheeled vehicles) and the AN/VVS-2 Driver's Viewer (tracked vehicles) as the main vision system for night driving operations. Over the past few years, the fielding of the Driver's Vision Enhancer (DVE) system has enabled the use of a Long-Wave Infrared (LWIR) Forward-Looking Infrared (FLIR) camera as a substitute to these Visible/Near-Infrared (V/NIR) image intensifier devices. The DVE FLIR has a 40-degree (horizontal) x 30-degree (vertical) Field of View (FOV). The driver views the sensor imagery on a 10.4" diagonal, Active Matrix Liquid Crystal Display (AMLCD) mounted approximately 10-14 inches in front of the driver's face. On wheeled vehicles, the sensor is mounted outside the vehicle on the roof, while the display is mounted inside the vehicle in front of the driver's face as shown in Figure 1.

With adequate training, DVE drivers can use the system to operate their vehicles through adverse environments and road conditions. FLIR sensors provide an advantage over image-intensified sensors for seeing through fog or smoke, seeing dirt paths on a cluttered forest road, and hot objects such as trees, rocks, or people. The system, however, does possess some man-machine limitations and is not a 24-hour, all-weather sensor solution; it can yield marginal imagery during thermal crossover periods and in wet environments. The sensor head is slow to rotate, and requires the driver to remove one hand from the steering wheel to pan/rotate the sensor head. DVE operators can experience eye fatigue from looking at an un-stabilized flat-panel display mounted at short viewing distances.

In June 1997, the US Army CERDEC Night Vision & Electronic Sensors Directorate (NVESD) and Kaiser Electronics began a cooperative program to develop a Head-Tracked Vision System (HTVS) to address the man-machine limitations of the DVE for driving applications. The HTVS program has been jointly funded by NVESD, the Project Manager for

Night Vision, Reconnaissance, Surveillance, and Target Acquisition (PM-NV/RSTA), and Kaiser Electronics. The initial goal of the HTVS program was to investigate the advantages of a head-tracked, head-displayed system for driving. Shortly into the program, the goals were expanded to include the examination of dual-waveband sensors and additive image fusion, for both increased driving capability and the possibility of near 24-hour operations. In July 2001, the program further expanded to include the evaluation of the feature-level image fusion algorithms developed by Sarnoff Corporation. The HTVS program has evolved into a mature vision system that could be a candidate for future thermal OMNIBUS DVE solicitations and future combat vehicles.

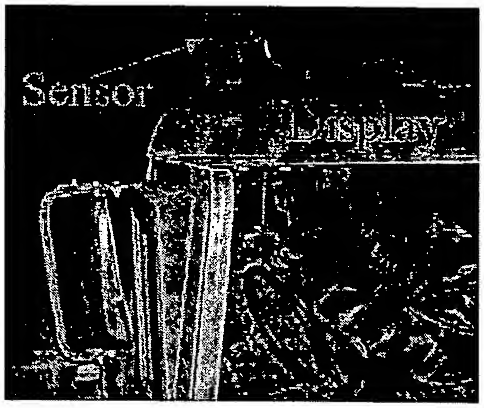

Figure 1.a Pre-Production DVE on wheeled vehicle.

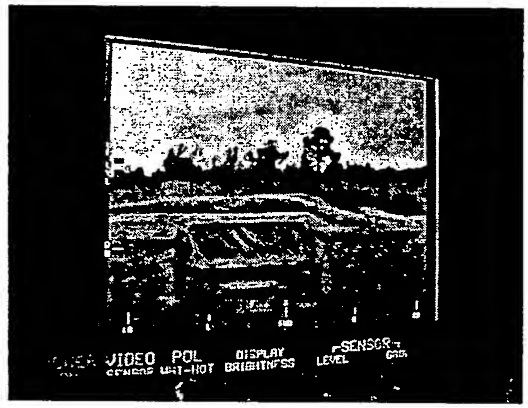

Figure 1.b Video imagery from DVE display.

A previous paper discussed in detail the design / performance parameters of the HTVS components / system, and the installation of the system on wheeled and tracked vehicles [1]. A second paper provided the results of a pilot study to assess the advantages of the HTVS vs. the DVE vs. Night Vision Goggles for driving applications, and provided an introductory discussion of the advantages of additive image fusion [2]. A third paper described ongoing improvements to the HTVS components, and provided an introductory discussion of additive versus feature-level image fusion in the HTVS [3]. This paper will focus on the recent field exercises and image analysis efforts to characterize the advantages of image fusion over single-waveband imagery, and to characterize the respective strengths/weaknesses of additive image fusion versus feature-level image fusion.

2. HTVS BASELINE DESIGN

The HTVS consists of four major components: the Head/Helmet-Mounted Display (HMD), the control unit (or system computer), the gimbal, and the optical tracker unit, as shown in Figure 2. The gimbal is mounted on the outside of the vehicle and acts as the pan and tilt mechanism for the FLIR and Image-Intensified CCD (IICCD or "IITV") sensors. The driver wears the HMD to view the imagery from the sensors. This HMD is head-tracked and slews the gimbal. The sensor gimbal duplicates the pan and tilt movements of the HMD. The head-tracker is a hybrid inertial and optical head-tracker. The inertial component tracks the head movements of the user both inside and outside the vehicle. This inertial tracker, however, requires periodic re-calibration, which is performed by the optical tracker. The control unit is the heart of the system. It performs all of the head-tracker calculations and controls the gimbal position. The control unit also includes video processing and performs the additive image fusion.

2.1 Baseline Helmet-Mounted Display (HMD)

The baseline HTVS HMD mounts on the infantryman's PASGT helmet as shown in Figure 2, or the combat vehicle crewman's helmet (or "CVCC", not shown). The HMD has been designed to accommodate all helmet sizes and requires no tools for installation. The HMD consists of two display oculars. Each ocular has a single-prism optic providing a 40 x 30 degree FOV with approximately 25 mm eye relief. This 40 x 30 degree FOV matches the sensors' FOV to maintain unity magnification and prevent motion sickness caused by mismatched fields of view. Each eye has a 0.9 inch, 1024 x

768 monochrome green AMLCD display. Each ocular has inter-pupillary distance (IPD) adjustment and rotational adjustment for optimal viewing of the display. The fore/aft position of both oculars can also be adjusted via a knob above the oculars. The head-tracker module is mounted on the front side of the PASGT helmet facing the windshield.

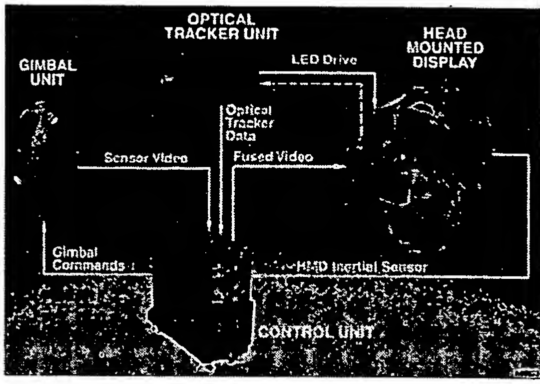

Figure 2.a HTVS diagram

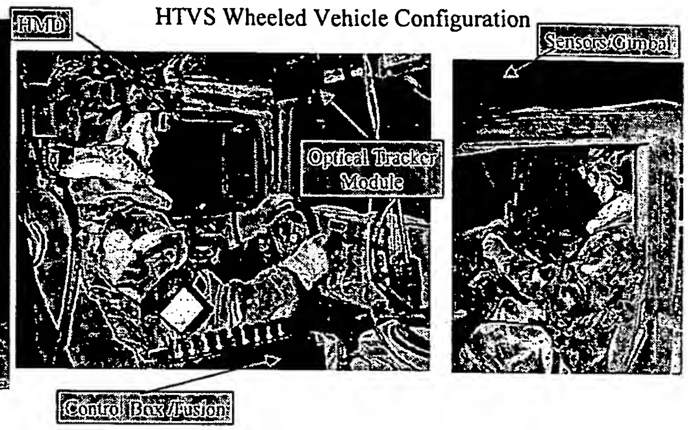

Figure 2.b HTVS installed on High Mobility Multipurpose Wheeled Vehicle (HMMWV).

2.2 Baseline head tracker

The baseline head-tracker is a hybrid inertial and optical tracking system. An inertial tracker cube with embedded rate sensors is mounted on the HMD to measure angular rates of rotation and linear acceleration along three perpendicular axes, which are then converted to the head movements of yaw, pitch, and roll [4]. A second inertial cube is mounted in the vehicle (housed in the optical tracker unit) to act as a vehicle motion reference. The inertial tracking system operates upon the differences between the data of the two cubes to calculate the corrected head position with respect to the vehicle motion [5]. The inertial tracking system tracks angular motion at a maximum rate of $\pm$ 1,000 degrees/second. The inertial tracker will operate with the HMD both inside and outside the vehicle. The rate sensors in the inertial cubes drift over time and require a periodic external re-calibration. The optical tracker corrects for this drift.

The optical tracker consists of three LEDs mounted in a triangle pattern on the HMD as shown in Figure 2.a. Two Photo-Sensor Detectors (PSDs) are mounted in the optical tracker unit and track the motion of the LEDs. Each LED is pulsed at a specific time and the PSDs sense the amplitude of each LED. The LED amplitude measured by the PSDs varies as a function of the distance between the LED and the PSDs. Two PSDs are used to triangulate the position of each LED. The optical tracker unit in the wheeled-vehicle configuration (HMD on the PASGT helmet) is mounted above the windshield facing the driver. The LEDs and the inertial cube on the HMD are mounted on the front side of the HMD.

The optical tracker will continuously correct the drift of the inertial tracker as long as the LEDs are in the "head box". However, the inertial-tracker drift only needs to be corrected every few minutes, which enables limited use of the HMD outside the vehicle.

2.3 Gimbal

The HTVS gimbal houses and acts as the pan and tilt mechanism for the FLIR and IITV sensors. The gimbal is 11 inches high and 7 inches in diameter as illustrated in Figure 2.a. The payload ball is 7 inches high and 4 inches wide. The gimbal with the two-sensor payload weighs approximately 15 lbs. The gimbal requires a nominal 24-Volt DC power (16-28 VDC range) and receives position data via an RS-232 serial link from the HTVS control module. The gimbal can continuously rotate 360 degrees and has $\pm$90 degrees elevation range of movement. In both azimuth and elevation, the maximum slew rate is 200 degrees/second and the peak acceleration is 1,000 degrees/sec^2. The gimbal has been sealed to withstand heavy rain, and has been designed to survive tree-branch strikes.

3. SENSORS AND ADDITIVE IMAGE FUSION

The HTVS sensor payload consists of LWIR (i.e., 8-12 micron waveband) and V/NIR (i.e., 0.4-0.9 micron waveband) cameras. These sensors are vertically mounted in the gimbal to minimize horizontal parallax. Both sensors have a 40-degree (horizontal) x 30-degree (vertical) field of view, but these respective fields of view have not been matched to the level required for pixel registration of the two sensors. It is noted that the FLIR sensor provides 8 pixels per degree, while the V/NIR sensor provides over 19 effective pixels per degree.

The FLIR sensor consists of an uncooled micro-bolometer possessing a detector noise-equivalent temperature difference (NETD) of approximately 80mK over its 320 x 240 array of 2-mil pixels. This sensor employs a custom objective lens having a focal length of 22mm and an F-number of 1.0. The FLIR electronics' digital signal processing allows for dead detector-pixel substitution, automatic level and gain adjustment via an Automatic Gain Control (AGC), and image histogram optimization. The RS-232 serial control also allows for manual control of the sensor level, gain, and polarity (i.e., white-hot vs. black-hot imagery). The video output is analog RS-170.

The image-intensified CCD sensor employs an 18mm Generation 3 image intensifier tube that is fiber-optically coupled to a CCD camera. The image intensifier tube has a limiting resolution of 64 lp/mm. An auto-gating circuit in the tube power supply enables both day and night operation. During daytime operation the gain is reduced via duty-cycle gating to prevent damage to the image intensifier tube. During nighttime operations the gain is raised to maximum levels for optimum nighttime viewing. The CCD camera employs a commercial 768 (horizontal) x 494 (vertical) pixel array in a 2/3-inch format. The video output is again analog RS-170. The IITV utilizes a commercial objective lens having a focal length of 17mm and an F-number of 1.4.

The FLIR sensor and the IITV have complementary strengths. The FLIR detects temperature differences in a scene and is not affected by ambient illumination. FLIR sensors are very good for seeing hot targets in a busy background, seeing through fog, and seeing paths through a cluttered forest (i.e., the dirt path will be a different temperature than the leaves or brush around the path). Uncooled LWIR FLIRs, however, are not as useful during thermal-crossover periods at night or after prolonged periods of rain, and cannot image cultural lighting, laser/LED lighting, and text on street signs.

IITVs amplify ambient illumination of the scene. IITVs are very good for seeing under moonlight conditions and enable imaging of moonlight shadows, headlights, flashlights, cultural lighting, laser/LED lighting, and text on road signs. IITVs produce high-resolution images with greater texture and detail than uncooled FLIR sensors. IITV imagery is degraded, however, in fog and under thick forest canopies with little ambient lighting. Underbrush and scrub vegetation can often make it difficult to discriminate between a forest path and its surroundings.

Many parties have noted the potential synergy that could result from a suitable combination of the imagery from LWIR and V/NIR sensors. For example, daylight hyperspectral measurements have indicated that the V/NIR and LWIR wavebands are highly uncorrelated [6]. Many disparate analyses and field experiences indicate that for optimal mobility operations, the ability to view scene information from *both* sensors is required. For this purpose, additive image fusion was initially incorporated into the HTVS video circuitry. This video circuitry takes the analog video outputs from each sensor and adds the two GEN-locked video streams together to form one image, which is displayed on the HMD. The user has the ability to adjust the fusion controller from all FLIR (100% FLIR and 0% IITV) to all IITV (0% FLIR and 100% IITV), or any mix of the two sensors (e.g., 70% FLIR / 30% IITV or 35% FLIR / 65% IITV).

Given the minimal extant level of field experience relating to image fusion for night mobility at this early program point, relatively few system design concessions were made to optimize image fusion. The primary consideration was to mount the sensors vertically to avoid horizontal parallax, and to minimize the residual vertical sensor separation to less than 3 inches. It is noted that for ground mobility operations, there is no foolproof way to compensate for sensor parallax. In aerial reconnaissance operations, a ground plane can be fitted to a down-looking field of view to compute relative distances to various objects in the FOV. In ground off-road operations, however, there is no way short of sophisticated motion parallax processing to determine the distance of a given object on the basis of its particular position in the field of view. Thus it is important to design the system from the outset to minimize image disparity due to parallax. The HTVS

has been designed to have no intrinsic horizontal sensor parallax, and its vertical parallax is insignificant (i.e., less than the IITV optimum resolution) at distances greater than about 75 meters.

The primary fusion design limitation of the HTVS is presently the lack of pixel-level registration between the two sensors. As noted previously, the IITV objective lens is a commercial unit that matches the FLIR FOV within about a degree, which is about an order of magnitude greater than that required for sensor registration. Optical methods for achieving sensor registration are generally insufficient in any case, because they require custom optics manufactured to extremely tight tolerances. For example, a tight tolerance (1%) for the focal length of the IITV objective lens could still result in about a 3-pixel discrepancy at 15-degrees off axis for even the IITV's modest array, and thermal/physical tolerance effects could only increase this discrepancy.

Sensor registration, however, could be quite readily achieved and maintained by periodic image processing correction, which would adjust the FOV mapping of one or both sensors via look-up tables. Such sensor registration techniques can be straightforwardly inserted into the HTVS whenever program resources permit. Until this is achieved, the additive fusion images produced by the HTVS should be assessed mainly for moderate-scale image features and overall scene contrast, especially in the outer portions of the FOV, where pixel misregistration is the greatest.

4. ADDITIVE VS. FEATURE-LEVEL IMAGE FUSION EVALUATION

NVESD has conducted a series of field exercises to better characterize the benefits of additive and feature-level fusion over single-sensor imagery in the HTVS. For all imaging reported herein, the thermal sensor was operated with automatic gain/level control, and its polarity was equally split between black-hot and white-hot.

Additive fusion, as discussed in section 3 above, employs a single (but user adjustable) fractional weighting for all the features of each sensor's image. In July 2001, NVESD and Sarnoff Corporation began a cooperative effort to refine Sarnoff's "feature-level" multi-resolution (pyramid) algorithms for image fusion. This approach employs digital processing techniques to select ***at each image point*** only the sensor with the strongest modulation (contrast), and to utilize only those features to reconstruct the fused video image. This selection process is performed simultaneously at multiple scales of the image, which are then combined to form the reconstructed fused image. The sub-weighting of each image scale can be individually modified, such that large-scale contrast or small-scale features can be selectively enhanced. The primary objective of the NVESD/Sarnoff collaboration is to determine if one can develop a robust weighting scheme (or at most a few such schemes) for each of the image scales that would be effective for the majority of nighttime field scenarios. An important adjunct to the Sarnoff image processing is a GUI-based function that specifically minimizes the sensor misregistration, as noted above. This alignment function consists of an affine transform, which includes translation, rotation, scale, and skew. More detail on the Sarnoff fusion algorithm is provided in the following subsection.

4.1 Sarnoff Corporation's feature-level image processing

Sarnoff has developed feature-level fusion technology over many years. Sarnoff's feature-level fusion is based on multi-scale (multi-resolution or pyramid) image processing algorithms providing pixel-level selection across multiple image scales. Sarnoff's algorithms are often referred to as pattern-selective (or Laplacian pyramid) image fusion, and have been implemented in real-time systems with the Acadia chip on a single PCI board, and can ultimately be implemented in a single chip with low latency.

4.1.1 Fusion algorithm

Sarnoff's feature-level image fusion [7,8] is implemented within a multi-resolution pyramid (or wavelet) image representation. Each source image (FLIR and IITV) is first transformed into a Laplacian pyramid representation. This has the effect of decomposing the image into local pattern structure of different scales. At each image position and scale, that source which has the highest-contrast feature content is then selected for inclusion in a corresponding pyramid representation for the fused image. The fused image itself is recovered through an inverse pyramid transform.

Figure 3 shows the two standard forms of the pyramid used in fusion: the Gaussian (or low pass), and the Laplacian (or band pass) pyramids. The base-level Gaussian pyramid, top left, is just the original image. Each successive Gaussian level is reduced in resolution and size by a factor of two, and is obtained by low pass filtering and subsampling the prior level. Each level of the Laplacian pyramid represents the difference between the corresponding and next lower resolution Gaussian levels. It is obtained by applying an appropriate band-pass filter to the Gaussian levels. Note that each level of the Laplacian highlights local edge or feature structure in the source image, and that different levels highlight structure at different scales. A key property of the representation is that these levels can be added together to recover the original image.

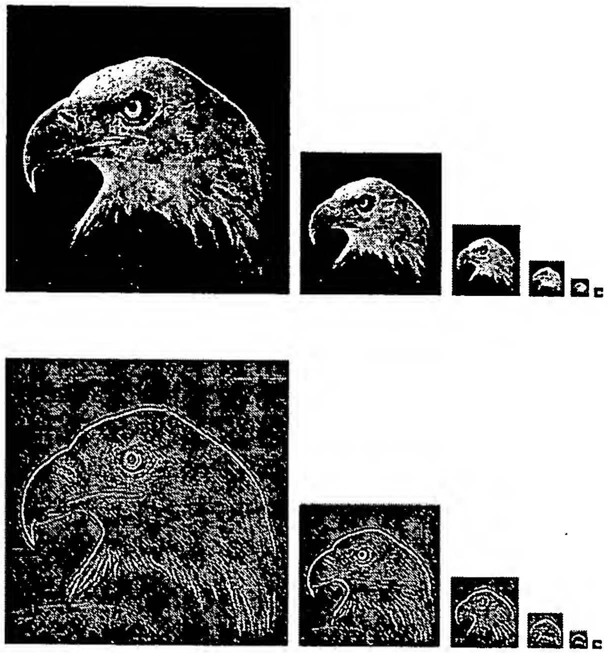

Figure 3. Gaussian low-pass pyramid (above) and Laplacian band-pass pyramid (below). The Laplacian highlights image features at multiple resolutions, and provides a framework for feature-level image fusion.

In the feature-level fusion approach a Laplacian pyramid structure is built for each source image. A pyramid for the fused image is assembled from the source image pyramids: the value assigned to each sample of the composite pyramid is just that source image sample at the corresponding image location (i.e., pyramid level and x-y position) that has the largest magnitude. This "select max" rule is illustrated in Figure 4. The fused image is then obtained from the composite pyramid through an inverse pyramid transform. The select-max process collects the highest contrast features for inclusion in the fused image. The inverse pyramid transform then merges these features into a single coherent image. The processing steps required for feature-level fusion can be performed continuously, in real time, using Sarnoff's Acadia image processing chip [9].

4.1.2 Tuning feature-level fusion

Use of the Laplacian pyramid framework in fusion provides a convenient means for controlling the process to best match the source images. For example, a basic means for image enhancement is "spectrum specification". The technique enhances certain spatial frequency bands of an image while reducing others. Images are made sharper, for example, by increasing the amplitude of high spatial frequencies while reducing the amplitudes of low spatial frequencies. This process is normally implemented within the Fourier-transform domain. It can also be implemented within the Laplacian pyramid by simply scaling each pyramid level by an appropriate factor prior to image reconstruction.

The feature selection process can also be biased to favor one source image or the other simply by weighting source image values during the selection process. Fusion can be tuned to characteristics of the image by using different bias factors on each pyramid level. For example, the process may be biased to favor the FLIR over the IITV at higher spatial frequencies when IITV noise is a problem, and to favor the IITV over the FLIR at lower spatial frequencies, where IITV noise is less prominent.

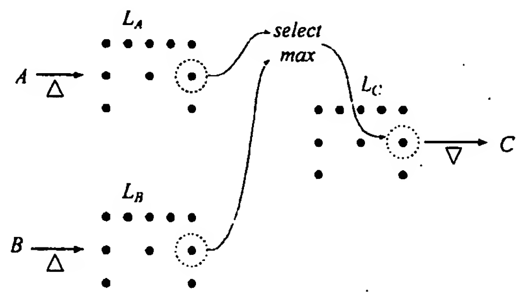

Figure 4. Schematic example of the feature-level fusion of source images *A* and *B* to form the composite image *C*. Dots represent individual samples in the respective pyramid structures.

4.2 Fusion imagery from nighttime field exercises

The images for this paper were generated during a field exercise conducted on 31 Jan 02 at a suburban location in Virginia. Images from one scene in this exercise will be examined from both qualitative and analytical standpoints. For this exercise, we not only recorded single-sensor imagery, but also recorded in real time the additive fusion imagery employed by the HTVS operator. We also recorded all three types of data while the HMMWV platform was in motion, and the HTVS was being actively scanned. The feature-level fusion images, in contrast, were generated by Sarnoff Corporation afterwards on selected still frames.

The weather conditions for this field exercise were cool (50s °F) and misty, with an intermittent mist/drizzle that significantly reduced scene temperature differences for the thermal sensor. The moon was below the horizon, so the ambient illumination levels were determined by the weather conditions and the cultural lighting in the vicinity, probably resulting in a high extreme of no-moon illumination. Two scenes from this exercise will be treated below: (1) An asphalt road under a diffuse hardwood forest canopy; and (2) An open meadow with high grasses and cattails.

The asphalt road images in Figure 5 below illustrate important aspects in which the fused image can provide improved situational awareness over either single-sensor image. The scene consisted of two persons standing over 100m down the asphalt road. Note that the two persons (one kneeling, one standing) can be easily discerned in the black-hot thermal sensor's image, though they are effectively imperceptible in the IITV image. The road striping and sky/tree-line interface, in contrast, can only be seen in the IITV image. Moreover, the IITV image clearly depicts a battery-powered NIR LED that had been placed on the road 20-30 meters away, although it is only an inconspicuous black spot in the thermal sensor's image due to its warm 9V battery. The additive fusion image clearly depicts all these important scene features. This additively fused image was generated according to a 70% IITV / 30% thermal sensor weighting scheme.

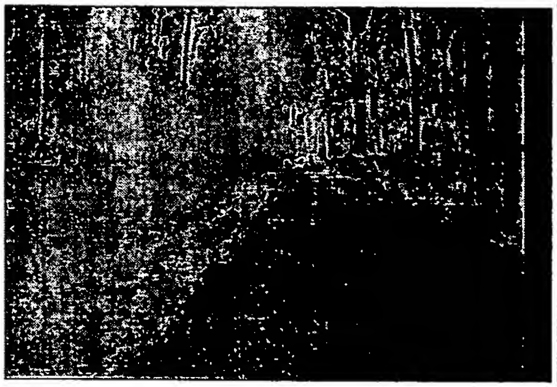

Figure 5.a Thermal Sensor (black hot)

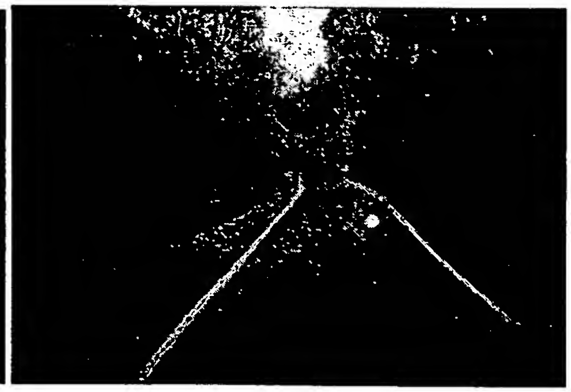

Figure 5.b IITV

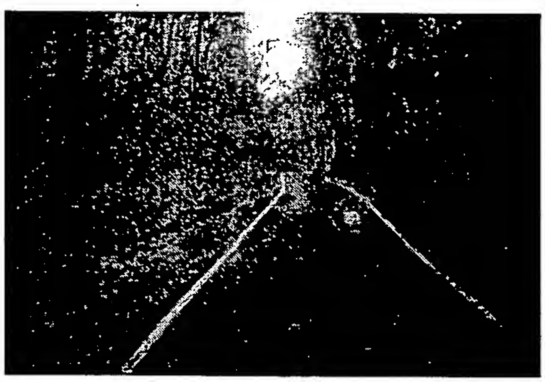

Figure 5.c Histogram-Optimized Additive Fusion

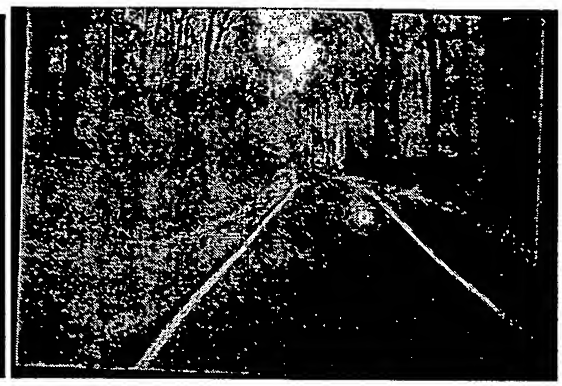

Figure 5.d Sarnoff Feature-Level Fusion

Figure 5. Single-Sensor vs. Fusion Images for Asphalt Road Scene

The feature-level fused image shows clearly improved feature resolution and general contrast over the additively fused image. Sarnoff's algorithms for sensor registration and for selective boosting of given spatial frequency bands are probably significant contributors to these improvements. On the other hand, the feature-level fusion approach also preserves more IITV noise in the fused image, while the additive fusion reduces the contrast of the IITV noise. These effects result from fundamental differences between the two fusion approaches, and are treated later in more detail.

The second scene was located in a meadow with high grasses and cattails, and was bisected by a dirt road. Weather and scene illumination conditions were the same as for the previous scene. The images from this scene are provided in Figure 6 below. This scene consists of two persons on a dirt path, with a pick-up truck and a cinder-block building in the background. The person on the right is holding a field radio with an LED indicator light in his left hand, and a weapon in his right hand. The combined effect of the subdued thermal signatures and the large thermal targets has forced the white-hot thermal sensor's AGC to effectively "black out" nearly everything else in the scene, except for the truck, building, and distant tree trunks. The IITV image clearly indicates the brush texture to either side of the path, along with the men, vehicle, and building. The respective sensors, however, provide reversed contrast for many of these salient scene features (note right man's hat, truck grill, building windows). The sky, though not depicted here, would also present reversed contrast in the two sensors' images. Our general experience has been that the thermal sensor's black-hot mode is the most suitable polarity for additive fusion with V/NIR imagery. An 80% IITV / 20% thermal sensor weighting scheme was employed for the additive fusion image in Figure 6.

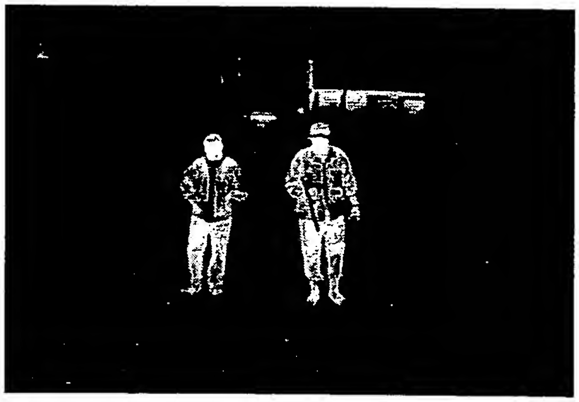

Figure 6.a Thermal Sensor (white hot)

Figure 6.b IITV

Figure 6.c Histogram-Optimized Additive Fusion

Figure 6.d Sarnoff Feature-Level Fusion

Figure 6. Single-Sensor vs. Fusion Images for Meadow Scene

The asphalt road images in Figure 5 were digitized and subjected to histogram analysis. A first-order qualitative interpretation of the single-sensor images is that the thermal sensor image has good contrast but poor gray scale, while the IITV has better gray scale but worse contrast and noise. This interpretation is supported by the analysis of the image histograms in Figures 7.a and 7.b below, which present the single-sensor and fusion histograms for the entire image.

Examination of Figure 7.a readily shows that the thermal image has a large number of very bright pixels (i.e., above gray level #175) and has essentially no dark pixels (i.e., gray levels #0 - 50). Visual inspection of the image confirms that nearly everything in the scene except for the men, road, and tree trunks approaches saturation in the thermal image. Figure 7.a indicates that the IITV image has a much broader gray scale distribution.

Visual inspection confirms the IITV image's generally good gray scale distribution, with nearly saturated pixels only in the lighted/sky areas and dark/black pixels only at the image corners. The IITV image's dark corners resulted from vignetting of its objective lens. This particular lens was a non-optimum commercial unit, which has been utilized on an interim basis until a better lens can be employed. The additively fused image has both a lower incidence of nearly saturated pixels than the thermal sensor's image, and a broader gray scale distribution than either sensor's image.

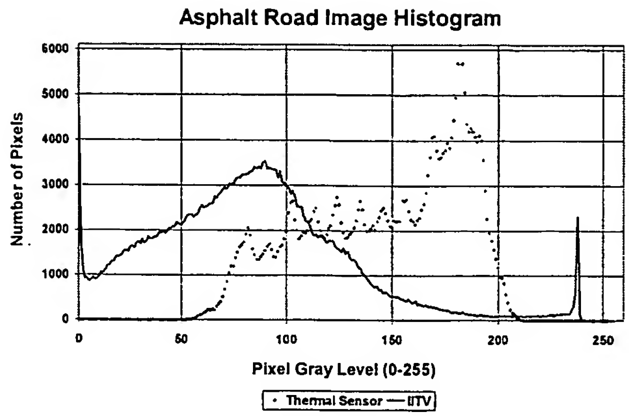

Figure 7.a Individual Sensor Histograms

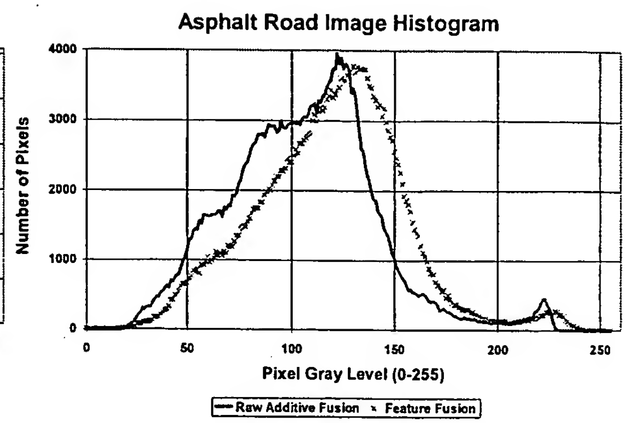

Figure 7.b Fusion Histograms

Figure 7. Single Sensor, Fusion Histograms for Asphalt Road Scene

Figure 7.b illustrates a basic shortcoming of "raw" additive fusion: it yields generally lower *overall* image contrast than that of either individual sensor. The reason for this is fundamental. Consider two given features in the asphalt road scene: the two men, and the tree line above them in the background. The two men have very high contrast in the thermal sensor's image, but are effectively invisible in the IITV image. They are readily discernible in the additively fused image, but have lower contrast than in the thermal sensor's image, because of the negligible-contrast contribution of the IITV to these features. The reverse situation holds for the tree line in the background. It is readily discernible in the fused image, but it has lower contrast than in the IITV image, due to the low-contrast contribution of the thermal sensor to this feature. When this principle is applied across the entire fused image, it can be considered as effectively adding a "DC" component to the contrast of most spatial features, thereby lowering overall scene contrast. Raw additive fusion shows *more* features than either sensor, but the contrast of these features is almost always lower than that in the lead sensor for a given feature. The only exception is when a given scene feature has the same contrast in both sensors. Additive fusion would not reduce the feature's contrast in this case, but this is a superfluous condition, because additive fusion would also provide no benefit over either individual sensor's image.

Sarnoff's feature-level image fusion, in contrast, does not combine both sensors' inputs at each image point; instead, the algorithm utilizes *only* the sensor with the highest modulation (contrast) at each image point. This approach effectively obviates the "DC" pedestal problem with additive fusion, and enables generally higher image contrast. Sarnoff's feature-level algorithm, however, also tends to preserve the contrast of the IITV noise, since it cannot yet effectively discriminate between the noise and feature information generated by the IITV. It consequently injects an undesirable amount of IITV noise into the fused imagery at the low end of ambient illuminations (e.g., clear and shadowed no-moon conditions). Some of this effect can be reduced by using feature-level tuning as described above in section 4.1.2.

Examination of Figure 7.b corroborates the above observations. In the additively fused image, essentially no pixels are found in the lowest 20 levels of the 256 total gray levels. This "black" pedestal could be entirely removed with very little loss of scene detail. Although the Sarnoff feature-level fused image shows a similar black pedestal, it has a generally better gray scale distribution, with less gray-value clustering than that found in the additively fused image.

The HTVS does not presently perform histogram optimization of *fused* imagery, even though such a basic functionality has been provided for the FLIR imagery. The HTVS developer is ultimately planning to incorporate such a functionality, however, and NVESD is currently evaluating the benefits of various schemes for histogram optimization. The result of one such scheme for optimizing additive fusion imagery is illustrated in Figure 8. This approach utilizes a relatively simple algorithm for more effectively mapping the image gray levels into the available dynamic range of the display (i.e., 8 bits).

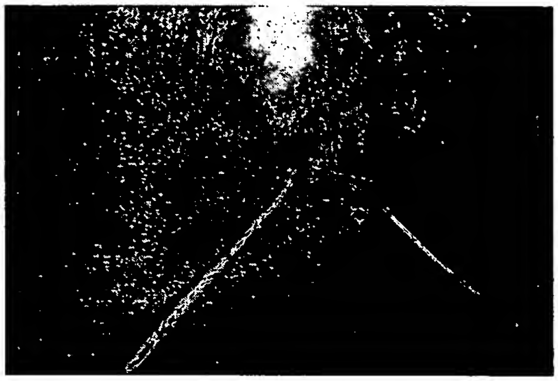

Figure 8.a "Raw" Additive Fusion of Asphalt Road Scene

Figure 8.b Histogram-Optimized Version of Image

5. CONCLUSIONS

From our initial HTVS field exercises and image analysis, the following general observations have emerged as consistent themes:

1. Sensor registration is of paramount importance if fused imagery is to retain the optimum resolution characteristics of the lead sensor. Horizontal parallax should be eliminated via the system design, and vertical parallax should be minimized. Sensor fields of view should also be optically matched as closely as possible, but this is probably not sufficient in itself, due to environmental effects and simple wear/tear from off-road usage. A periodic software-based sensor mapping adjustment will probably also be required. Sarnoff Corporation already incorporates this algorithm as part of their standard image processing.

2. In clear and shadowed no-moon conditions, the IITV imagery has significant signal-related noise that should ideally not be translated into the fused imagery. Additive fusion inherently suppresses such noise, but to only a modest degree. The Sarnoff fusion processing needs to be supplemented by digital algorithms that minimize the contribution of this noise to the final fused image.

3. "Raw" additive fusion typically results in imagery that has generally lower contrast than either of its constituent sensors. This effect can be mitigated by one or more histogram optimization schemes, which potentially incur little processing power, system real estate, or image latency.

4. In head-slewed mobility systems such as the HTVS, image latency must be minimized. Many sources indicate that the maximum tolerable image latency is in the 40 - 60 millisecond range [10]. Additive image fusion can be designed to have negligible added image latency. The Sarnoff image fusion implementation presently incurs either a 2-field or 2-frame added latency, in part due to the need to have a complete image in the processing buffer to start the operation. Sarnoff is presently working on ways to minimize this added processing latency.

5. Corresponding V/NIR and LWIR images appear to have less anti-correlation (i.e., reversed contrast) when the thermal imager is in black-hot polarity. This polarity therefore appears to yield more consistent fusion results than the white-hot polarity for mobility (i.e., non-targeting) applications.

6. A surprising observation is that typical additive fusion weightings have clustered around a 65% IITV / 35% Thermal Sensor balance. We speculate that the low/moderate-contrast scene texture provided by the IITV requires a higher relative weighting to be adequately translated into the fused image. The high-contrast thermal imager features, however, can still be adequately expressed in the fused image despite a lower relative weighting.

ACKNOWLEDGMENTS

The authors would like to express their appreciation for the support from PM-NV/RSTA and PM-FLIR. They would also like to thank the Kaiser Electronics team for their devotion and hard work over the past five years in developing the HTVS, especially Richard Reed, Curt Casey, and Terry Harper. We would also like to thank the Sarnoff Corporation's image processing group, including Peter Burt and Mike Hansen, and NVESD's Steve Hart and Russell Draper for their support during the HTVS field exercises.

REFERENCES

1. C. Reese, E. Bender, "Multispectral image-fused head-tracked vision system (HTVS) for driving applications," Proc. SPIE, 4361, pp. 1-11, 2001.

2. C. Reese, E. Bender, R. Reed, "The Use of an Image-Fused, Head Tracked Vision System (HTVS) for Driving Applications," Proc. National Military Sensors Symposium Conference, November 2001.

3. C. Reese, E. Bender, R. Reed, "Advancements of the head-tracked vision system (HTVS)," Proc. SPIE AeroSense, April 2002.

4. C. Casey, "Helmet-mounted displays on the modern battlefield," Proc. SPIE, 3689, pp. 270-277, 1999.

5. E. Foxlin, "Head tracking relative to a moving vehicle or simulator platform using differential inertial sensors," Proc. SPIE, 4021, pp.133-144, 2000.

6. S. Horn et al, "Fused reflected/emitted light sensors," Proc. SPIE AeroSense, 4369, pp. 1-13, 2001.

7. P.J. Burt, "Pattern selective fusion of IR and visible images using pyramid transforms," National Symposium on Sensor Fusion, 1992.

8. P.J. Burt and R.J. Kolzcynski, "Enhanced Image Capture Through Fusion," Proc. IEEE Int. Conf. on Computer Vision, 1993.

9. G.S. van der Wal, M.W. Hansen, M.R. Piacentino, "The Acadia Vision Processor," Proc. IEEE Int. Workshop on Comp. Arch. for Machine Perception (CAMP), Italy, September 2000, pp. 31-40.

10. Biberman (editor), *Electro-Optical Imaging: System Performance and Modeling*, pp. 26-27 – 26-35, Ontar Corporation/SPIE Press, Bellingham, WA, 2000.